
Introduction

COSPEDSpec is a python based tool for computing species tree from a set of incongruent gene trees:

1. First our earlier method COSPEDTree is applied as its first stage, to generate a possibly non-binary species tree S' covering all the input taxa.
2. Subsequently we apply a refinement on S', with reducing the number of extra lineages (deep coalescence) with respect to the input gene tree set, for producing a strict binary species tree.

Input source trees can be either in NEWICK format or in NEXUS format.

However, all the source trees should have identical input formats. They should be placed in a standard tree list file, according to the syntax of NEXUS or NEWICK formats. Such a tree list text file is to be provided as an input of this executable.

Output weighted supertree is generated in the NEWICK format.

Dependencies

COSPEDSpec is developed in Linux Systems (Ubuntu 14.04), using Python 2.7.

If a system already has python installed and the package has version lower than 2.7, then there can be problems, as tested. So, in such case, corresponding python package upgrade is requested.

The executable is tested with different versions of Ubuntu.

For systems having Ubuntu with lower versions, please notify in case of any errors.

The executable does not require any libraries to be installed prior execution.

Execution

Upon extracting the archive COSPEDSpec.zip, a folder named 'dist' is obtained. User needs to go inside to that folder and find that there is an executable named COSPEDSpec.

Check the permission of the executable – otherwise apply following command to make it an executable:

```
chmod +x COSPEDSpec
```

Note: If you want to copy the extracted archive to a different location, copy the entire folder, not only the executable only. Other files are supporting libraries and binaries associated with this executable.

Now you can execute COSPEDSpec with the following commands (assume the user is within the directory 'dist'):

EXAMPLE OF COMMANDS

```
./COSPEDSpec -I 'source_tree_input_filename' -p 'inp_file_format' -O 'out_species_tree'
```

Command descriptions:

1. Using -I command we specify the input filename (denoted by 'source_tree_input_filename'). User need to specify the absolute or relative path of the file containing the input gene tree dataset (maintained in a text file of standard tree list, in either nexus or newick format).
2. -p option is for specifying the input tree format (as denoted by 'inp_file_format'). If input file contains the trees in NEWICK format, then specify the option as (-p 1) (1 stands for newick). If input file contains the trees in NEXUS format, then specify the option as (-p 2) (2 stands for nexus). By default, p = 1 is set.
3. -O option is required for specification of the output filename ('out_species_tree') containing the species tree. The tree will be stored in Newick format, and the output file will be a simple text file. In addition, one text file will be created in the same directory. It will contain the timing information.

For any queries, please contact

Sourya Bhattacharyya
Department of Computer Science and Engineering
Indian Institute of Technology Kharagpur
<sourya.bhatta@gmail.com>